



Markov Chain Monte Carlo Posterior Density Approximation for a Groove-Dimensioning Purpose

José Ismael de La Rosa Vargas, Gilles Fleury, Sonia Esther Osuna, Marie-Eve
Davoust

► To cite this version:

José Ismael de La Rosa Vargas, Gilles Fleury, Sonia Esther Osuna, Marie-Eve Davoust. Markov Chain Monte Carlo Posterior Density Approximation for a Groove-Dimensioning Purpose. IEEE Transactions on Instrumentation and Measurement, 2006, Vol. 55 (N°1), pp. 112-122. 10.1109/TIM.2005.861495 . hal-00260575

HAL Id: hal-00260575

<https://hal-centralesupelec.archives-ouvertes.fr/hal-00260575>

Submitted on 5 Mar 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Markov Chain Monte Carlo Posterior Density Approximation for a Groove-Dimensioning Purpose

José I. De la Rosa, *Member, IEEE*, Gilles A. Fleury, Sonia E. Osuna, and Marie-Eve Davoust

Abstract—The purpose of this paper is to present a new approach for measurand uncertainty characterization. The Markov chain Monte Carlo (MCMC) is applied to measurand probability density function (pdf) estimation, which is considered as an inverse problem. The measurement characterization is driven by the pdf estimation in a nonlinear Gaussian framework with unknown variance and with limited observed data. These techniques are applied to a realistic measurand problem of groove dimensioning using remote field eddy current (RFEC) inspection. The application of resampling methods such as bootstrap and the perfect sampling for convergence diagnostics purposes gives large improvements in the accuracy of the MCMC estimates.

Index Terms—Gibbs sampling, indirect measurement, Markov chain Monte Carlo (MCMC), Metropolis–Hastings (M–H), nonlinear regression, perfect sampling, weighted bootstrap.

I. INTRODUCTION

IN MANY industrial applications, direct access to a measurand (m) is not possible; this is due to the inability to use transducers to measure m directly for any reason such as harsh environment, long distance, etc. Thus, the measurand process must be considered as an inverse problem [1], since the measurand estimation is needed. The characterization of all statistical knowledge upon this quantity of interest is naturally driven by the probability density function (pdf) $\varphi(m)$. Probabilistic inferences using Markov chain Monte Carlo (MCMC) methods are considered as another Monte Carlo simulation technique, and other measurand pdf estimation alternatives, by using a fully Bayesian framework [2]–[5]. The “Metropolis algorithm” has been used to solve difficult problems in statistical physics for over 40 years (1953). A generalization of this algorithm is introduced by “Metropolis–Hastings (M–H) algorithms” (1970). In the last few years, other related methods of “Gibbs sampling” have been applied to problems of statistical inference. The MCMC methods, such as Gibbs sampling and M–H algorithms, are powerful Markov-chain methods to simulate multivariate distributions and they have a real impact on Bayesian statistics [6], [7]. The M–H algorithms have been extensively used in physics and, more recently, exploited by statisticians [8].

The importance of Monte Carlo methods for inference problems in signal processing has grown in recent years (*Trans. on Signal Process.*, vol. 50, no. 2, Feb. 2002). This is due to the explosive increase in accessible computing power. Monte Carlo methods could be widely exploited, since one has the necessary computational resources, and these methods can significantly be used in a large class of problems addressed in practice. Monte Carlo methods have also a great degree of flexibility for the solution of challenging computational problems, such as optimization and integration. These kinds of problems abound in statistical signal processing. Literature [9], [10] show the potential usefulness of MCMC in signal processing. For the solution of our measurand problem, we have successfully used hybrid algorithms, firstly, an M–H algorithm for sampling from a complex likelihood density function, and secondly, the Gibbs algorithm for sampling from the posterior density.

An important problem in MCMC is the convergence surveillance of such methods. The convergence performance of MCMC can be improved by using a resampling scheme, e.g., the weighted bootstrap used in [11] (see also [12] and [13], which suggest a class of weighted-bootstrap techniques) and perfect-simulation [14] procedures (see also [15] and [16]). The final interest is to apply the MCMC methods in a realistic problem of indirect measurement (measurand estimation). The remainder of the paper is organized according to the following sections: Section II presents the general formulation of the problem of measurand estimation and the MCMC idea extended to the measurand uncertainty characterization. The Bayesian framework for parameter and measurand estimation is described in a more specified way in Section III, jointly with the classical MCMC uncertainty characterization. The analysis of convergence by resampling and perfect-sampling methods [coupling from the past (CFPP)] are briefly described in Section IV. A measurement complex problem of groove dimensioning using remote field eddy current (RFEC) inspection is given in Section V, and finally, some concluding remarks are given in Section VI.

II. MCMC FOR MEASURAND PDF ESTIMATION

The problem of pdf estimation for an indirect measurand is considered in this paper. This problem has been analyzed for a nonlinear Gaussian framework, and the results give the possibility to take up again the problem of pdf estimation in a more suitable or realistic framework (nonlinear Gaussian with unknown variance or non-Gaussian). In many applications, an unknown quantity m has to be estimated from a vector

Manuscript received March 14, 2005; revised October 7, 2005.

J. I. De la Rosa is with the Signal Processing Laboratory of the Engineering faculty, Universidad Autónoma de Zacatecas, Zacatecas, Mexico. He is also with the École Supérieure d'Électricité—Service des Mesures, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France (e-mail:ismaelrv@ieee.org).

G. Fleury, S. E. Osuna, and M.-E. Davoust are with the École Supérieure d'Électricité—Service des Mesures, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France (e-mail:Gilles.Fleury@supelec.fr).

Digital Object Identifier 10.1109/TIM.2005.861495

of observed values \mathbf{y} . This may be encountered in several domains such as nondestructive testing or the so-called indirect measurand [17], [18]. Measurement systems are formalized by two equations [1]:

- 1) the observation equation, which is described by the classical nonlinear regression model

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{e} \quad (1)$$

- 2) and the measurand equation (a nonlinear function of the parametric model)

$$m_\ell = g_\ell(\boldsymbol{\theta}), \quad \ell = 1, \dots, r, \quad \text{with} \quad \mathbf{m} = \{m_\ell\}_{\ell=1}^r \quad (2)$$

where \mathbf{y} represents the response (observed data, acquired data, or explained variables), \mathbf{x} is the vector of explicative variables or experimental protocol (for example, the measurand instants in an instrument, the sensor position, the frequencies used in the Eddy currents [17], etc.) of dimension p (or matrix $n \times p$), which is associated with data $y_i(y_i, x_i^\top)$, and $\boldsymbol{\theta} = \{\theta_j\}_{j=1}^p$ is the vector of functional parameters of dimension p , which will be estimated by an identification procedure. The errors model the system as a random process and they are given by \mathbf{e} . The set $\Theta \subset \mathbb{R}^p$ represents the admissible subset of parameters $\boldsymbol{\theta}$, and $\mathbb{X} \subset \mathbb{R}^n$ represents the admissible domain for the explicative variables. The measurand depends on the nonlinear mapping $m_\ell = \mathcal{G}_\ell(f)$.

We have already presented a methodology based on bootstrap techniques (see [19] and [20]) to obtain the pdf estimation of an indirect measurand. Currently, our work is focused on a Bayesian framework to obtain the posterior-pdf estimation of such a measurand, using MCMC methods. The idea is to obtain the posterior pdf

$$\varphi(\boldsymbol{\theta}|\mathbf{y}) \propto \varphi(\mathbf{y}|\boldsymbol{\theta})\varphi(\boldsymbol{\theta}). \quad (3)$$

The measurand depends on the nonlinear mapping $m_\ell = \mathcal{G}_\ell(f)$, and, given the parameters' posterior pdf, one may assume that the measurand posterior is obtained directly by the nonlinear mapping parameters' posterior pdf

$$\varphi(m_\ell|\boldsymbol{\theta}) = \varphi(g_\ell(\boldsymbol{\theta}|\mathbf{y})), \quad \ell = 1, \dots, r. \quad (4)$$

A complete diagram of the general problem is shown in Fig. 1, where some other complicated stages are depicted, such as the parameter estimation criterion $\mathcal{J}(\boldsymbol{\theta})$; for example, a classical criterion to be optimized is given by a functional of the errors

$$\mathcal{J}(\boldsymbol{\theta}) = \sum_{i=1}^n \psi(e_i) \quad (5)$$

the optimization procedure to be used, and the nonlinearity of the model $f(\cdot)$ retained for the measurand purpose [21], [22]. The propagation of statistical properties of the errors through

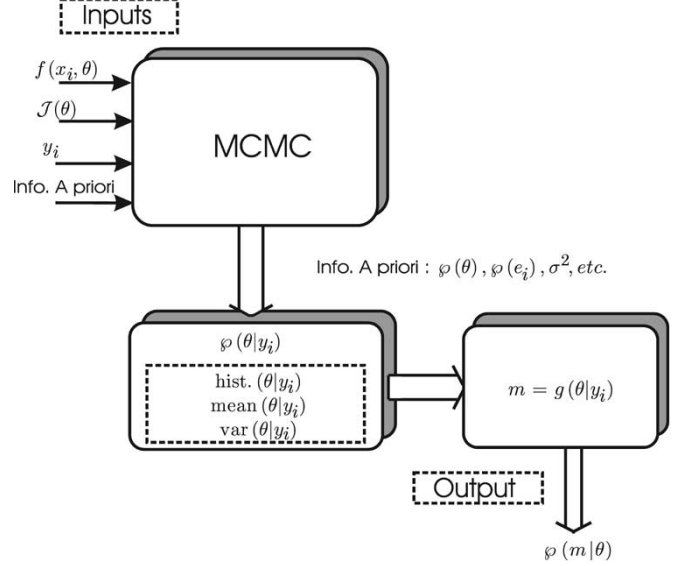


Fig. 1. Diagrams to illustrate the parameter and the measurand posterior pdf estimation using MCMC methods.

the observation equation permits to access statistical properties of model parameters and then propagate such statistical information through the measurand equation to obtain statistical properties of a measurand. The assumption upon the errors' probability law is important, and generally, a Gaussian assumption has been taken. Nevertheless, in real problems, where the variance is unknown or it is nonuniform (heteroscedasticity), a non-Gaussian assumption also could be more suitable, and the assumption of different laws such as uniform, approximately uniform, Laplace, mixture, etc., could be made. The assumption of an unknown law could also be considered, as assumed in applications of bootstrap methods.

III. BAYESIAN STATISTICS

The Bayesian framework is a good selection alternative to take advantage of all possible information known from complex systems (e.g., analysis of complicated statistical models). The advent of inexpensive high-speed computers and the simultaneous rapid development of stochastic integration methodology permit to exploit the use of MCMC methods in a Bayesian framework [6], [7]. The MCMC methods have been used in several Bayesian problems, e.g., analysis of models of varying size, multiple changepoints, and finite mixtures. Moreover, MCMC methods solves analytically intractable inference problems.

The interest of Bayesian statistical inference involves the task of estimation of complex distributions. It requires an additional input not required by frequentist procedures such as maximum likelihood (ML), called *a priori* probability distribution for the parameters $\varphi(\boldsymbol{\theta})$ and given the nonlinear mapping $m_\ell = \mathcal{G}_\ell(f)$, for the measurand $\varphi(m_\ell)$. The introduction of *a priori* distribution is the crucial element that converts statistical inference into an application of conditional probabilistic inference (taking advantage of prior information). For the consideration of combining *a priori* distribution for the parameters with the conditional distribution for the

observed data \mathbf{y} , let us get a joint distribution for all quantities related to the problem

$$\begin{aligned} \varphi(\theta_1, \dots, \theta_p, y_1, \dots, y_n) \\ &= \varphi(y_1, \dots, y_n | \theta_1, \dots, \theta_p) \varphi(\theta_1, \dots, \theta_p) \\ \varphi(\boldsymbol{\theta}, \mathbf{y}) \\ &= \prod_{i=1}^n \varphi(y_i | \boldsymbol{\theta}) \varphi(\boldsymbol{\theta}). \end{aligned} \quad (6)$$

From this, the Bayes' rule can be derived for the posterior distribution of the parameters, given the observed values \mathbf{y}

$$\varphi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\varphi(\boldsymbol{\theta}, \mathbf{y})}{\varphi(\mathbf{y})} = \frac{\prod_{i=1}^n \varphi(y_i | \boldsymbol{\theta}) \varphi(\boldsymbol{\theta})}{\int \prod_{i=1}^n \varphi(y_i | \boldsymbol{\theta}) \varphi(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (7)$$

The posterior distribution can also be expressed as a proportionality in terms of the likelihood

$$\varphi(\boldsymbol{\theta} | \mathbf{y}) = \mathcal{L}(\mathbf{y} | \boldsymbol{\theta}) \varphi(\boldsymbol{\theta}). \quad (8)$$

This shows how the introduction of a *a priori* distribution converts the expressions of relative plausibility contained in the likelihood into an actual probability distribution over parameter space $\Theta \subset \mathbb{R}^p$ and by the final nonlinear mapping over measurand and space $\mathbb{M} \subset \mathbb{R}^r$.

The final goal is therefore to obtain an accurate statistical characterization of the measurand quantity m_ℓ . Once the parameters posterior $(\boldsymbol{\theta} | \mathbf{y})$ is simulated, the measurand vector induced by the mapping $\mathcal{G}_\ell(f(\cdot))$ is given as

$$\widehat{m}_\ell = g_\ell(\boldsymbol{\theta} | \mathbf{y}), \quad \ell = 1, \dots, r. \quad (9)$$

The measurand pdf $\varphi(\mathbf{m})$ is then approximated by the empirical measurand pdf $\varphi(\widehat{\mathbf{m}})$, which is induced by $\widehat{\mathbf{m}}_1, \dots, \widehat{\mathbf{m}}_N$ using the different replications $(\widehat{\boldsymbol{\theta}} | \mathbf{y})_1, \dots, (\widehat{\boldsymbol{\theta}} | \mathbf{y})_N$ in the nonlinear mappings $\{\mathcal{G}_\ell(f(\cdot))\}_{\ell=1}^r$, taking advantage of the MCMC process (N is the number of simulations). Note that marginal measurand pdfs are also obtained, solving the integral

$$\varphi(\widehat{m}_\ell | \boldsymbol{\theta}) = \varphi(g_\ell(\boldsymbol{\theta} | \mathbf{y})) = \int_{\mathbb{M}} \varphi(\widehat{\mathbf{m}}) d\widehat{\mathbf{m}}_{-\ell} \quad (10)$$

where $d\widehat{\mathbf{m}}_{-\ell} = d\widehat{m}_1 \cdots d\widehat{m}_{\ell-1} d\widehat{m}_{\ell+1} \cdots d\widehat{m}_r$.

In cases where the parameter space has very high dimensionality, and the posterior distribution is very complex (for models where the dimension $p > 2$ and $r > 1$), obtaining a Monte Carlo estimate may then require the use of Markov-chain sampling methods. This obviously will happen in the evaluation of the integrals over parameter and measurand spaces (as will be seen in Section V), and it could be computationally very demanding. However, the Monte Carlo estimation could be successfully applied, since these methods only require calculation of probability densities for parameter values up to an unknown constant factor, and thus, it is not necessary to evaluate the integral in the denominator of (7).

IV. MCMC CONVERGENCE IMPROVEMENT

An important problem in MCMC is the convergence surveillance of such methods. We present two different strategies to accelerate the convergence of an MCMC algorithm. The term acceleration is used in numerical analysis to indicate hastening of convergence, diminishing the number of the chain transitions (t) (e.g., draws). The first approach is through resampling and the second one proposes a perfect-sampling scheme (see also exact sampling) based on CFTP procedures.

A. Weighted Bootstrap

A weighted-bootstrap scheme was introduced in [11]. The task is to compute a resample from the set $\boldsymbol{\theta}_{t,j}$, $j = 1, \dots, q$, independent identically distributed (i.i.d.) with $\varphi(\boldsymbol{\theta}_{t,j})$; the new sample is $\boldsymbol{\theta}_{t,j}^*$, $j = 1, \dots, q^*$, whose distribution $\varphi(\boldsymbol{\theta}_{t,j}^*)$ is closer to $\varphi(\boldsymbol{\theta})$ than $\varphi(\boldsymbol{\theta}_{t,j})$ is. Note that j indicates the size of a random vector of each one of the parameters $\boldsymbol{\theta} = [\theta_1, \dots, \theta_p]^\top$. The weights are given as follows:

$$w_j = \frac{\varphi(\boldsymbol{\theta}_j)}{\varphi(\boldsymbol{\theta}_{t,j})} \quad \text{and} \quad z_j = \frac{w_j}{\sum_{j=1}^q w_j}$$

where $\boldsymbol{\theta}_{t,j}$ are resampled according to the probabilities z_j . The resampled draw $\boldsymbol{\theta}_{t,j}^*$ will be applied to the next transition of the sampling algorithm to obtain $\boldsymbol{\theta}_{(t+1),j}$. Instead of using this approach, we have used a slightly different approach proposed for resampling the residuals $\widehat{r}_i = y_i - f(x_i, \boldsymbol{\theta}_t)$, where the random vector to be resampled now is of the same size as \mathbf{y} . The idea is the same; it also provides the same effect that the new resample gives: $\boldsymbol{\theta}_t^* \sim \varphi(\boldsymbol{\theta}) \rightarrow \boldsymbol{\theta}_{t,j}^* \sim \varphi(\boldsymbol{\theta})$. Such an approach was introduced by Shao [13] to approximate $\varphi(\boldsymbol{\theta})$ via nonparametric bootstrap simulation. We also assume that under regular conditions, $\Delta_t = \lim_{t \rightarrow \infty} \|\varphi(\boldsymbol{\theta}_t^*) - \varphi(\boldsymbol{\theta})\| \rightarrow 0$. Bootstrapping a modified residuals vector gives a weighted (or smoothed) version of residuals that can lead to a consistent estimator (e.g., weighted bootstrap)

$$\begin{aligned} \bar{r} &= \frac{1}{n} \sum_{i=1}^n \widehat{r}_i \\ \widehat{e}_i &= \frac{\sqrt{\frac{n}{n_q}} (\widehat{r}_i - \bar{r})}{\sqrt{1 - \frac{p}{n}}} \end{aligned} \quad (11)$$

where \widehat{e}_i corresponds to $\widehat{e}_i = y_i - f(x_i, \boldsymbol{\theta}_t^*)$. The procedure of bootstrapping recentered residuals (under symmetric assumptions of the error distribution) can be modified; in linear regression, the bootstrap sample size of \widehat{e}_i^* is modified ($i = 1, \dots, n_q$, such as $n_q < n$). However, in nonlinear regression, the n_q factor value is modified such as $n_q < n$ and satisfying $n_q/n \rightarrow 0$, but the bootstrap sample size remains fixed ($i = 1, \dots, n$). A restriction on n_q is that p/n_q should be reasonably small. For practical uses, n_q needs to be specified for a fixed n , and its optimal choice depends on the parameter-vector size. For example, one could begin with $n_q = n$ and then decrease this value till $n_q/n = 0.5$ if n is small, and $n_q/n \rightarrow 0$

if $n \rightarrow \infty$. The resampled draw \hat{e}_i^* produces indirectly the draw θ_t^* that will influence the next transition of the sampling algorithm to obtain $\theta_{(t+1)}$.

B. Perfect-Simulation Algorithms

The perfect-simulation techniques or exact sampling ideas were introduced by Propp and Wilson in 1996 [15]. The main task is to use CFTP and to use repeatedly the same sampler for generating upper and lower Markov chains started increasingly further back in time until the two chains are coupled at time 0 or coalesce when their transitions are determined by the same random numbers and then return the result as an exact simulation from a given target distribution. In the last works of Møller [14], the CFTP technique has been applied to improve convergence of conditional specified multivariate models such as autogamma and establishes the application of perfect simulation for an M–H algorithm for the same model.

1) *Gibbs Perfect Simulation*: Suppose that we want to make simulations from a target distribution $\varphi(\theta|\mathbf{y})$ with $\theta = [\theta_1, \dots, \theta_p]^\top$ being a p -dimensional discrete vector; suppose also that $\varphi(\theta|\theta_{-k})$, with $k = 1, \dots, p$ is the conditional distribution, where $\theta_{-k} = [\theta_1, \dots, \theta_{k-1}, \theta_{k+1}, \dots, \theta_p]$. Let $F_k(\cdot|\theta_{-k})$ denote the cumulative distribution function (CDF) of $\varphi(\theta|\theta_{-k})$ when the probability $P\{\theta_{-k}\} > 0$. We can then generate a Markov chain $\theta_{t,k} \in [\theta_{t,1}, \dots, \theta_{t,p}]$, $t = 0, 1, \dots$, started at $\theta_{0,k} = 0$ and using cyclic Gibbs sampling by setting

$$\theta_{t,k} = F_k^- \{R_{t,k} | \theta_{t,k \leftarrow}\}$$

where

$$R_{t,k} \sim \mathcal{U}(0, 1)$$

and

$$\theta_{t,k \leftarrow} = [\theta_{t,1}, \dots, \theta_{t,k-1}, \theta_{t-1,k+1}, \dots, \theta_{t-1,p}]$$

are the $p - 1$ states of the components just before the k th update at time t , and $F^{-1}(x) = \min\{y : F(y) \geq x\}$. We assume that $\theta_{t,k}$ follows the irreducibility property. Suppose now that $R_{t,k}$ are also defined back in time $t = -1, -2, \dots$; then, the CFTP is obtained by reusing these random numbers in the following construction of lower $L_t(n_t) = [L_{t,1}(n_t), \dots, L_{t,p}(n_t)]$ and upper $U_t(n_t) = [U_{t,1}(n_t), \dots, U_{t,p}(n_t)]$ processes, which are started at times n_t and generated forwards in time. Then, at each integer $n_t \in \mathbb{Z}$, put

$$\begin{aligned} L_n(n_t) &= 0 \\ L_{t,k}(n_t) &= F_k^- \{R_{t,k} | U_{t,k \leftarrow}(n_t)\}, \quad t > n_t \end{aligned} \quad (12)$$

and

$$\begin{aligned} U_n(n_t) &= D(n_t) \\ U_{t,k}(n_t) &= F_k^- \{R_{t,k} | L_{t,k \leftarrow}(n_t)\} \end{aligned} \quad (13)$$

with dominating chain $D(n_t) = [D_1(n_t), \dots, D_p(n_t)]$ given by the mutually independent components

$$D_k(n_t) = F_k^- \{R_{n_t,k} | \mathbf{0}_{-k}\}$$

and extending the definition of $F_k(\cdot|\theta_{-k})$ when probability $P\{\theta_{-k}\} = 0$. For $k \in \{1, \dots, p\}$ and $\theta_{-k} \in \mathbb{R}^{p-1}$, define

$$F_k(\cdot|\theta_{-k}) = \max \{F_k(\cdot|\beta_{-k}) : \beta_{-k} \leq \theta_{-k}, P\{\theta_{-k}\} > 0\}$$

which ensures that $F_k(\theta_k|\theta_{-k})$ is increasing in θ_{-k} . For the perfect-simulation algorithm, we use a strictly decreasing sequence of nonpositive starting times $n_t = n_j$ ($0 \geq n_1 > n_2 > \dots$), where $j = 1, 2, \dots$, and repeat to generate lower and upper processes $(L(n_t), U(n_t))$ until coupling happens at time $t = 0$, where we generate $(L_t(n_t), U_t(n_t))$, $t = n_t, \dots, 0$, until $L_0(n_t) = U_0(n_t)$; then $S = L_0(n_t)$, where S is the state selected as the perfect simulation from the target distribution $\varphi(\cdot)$. For application cases, it has been suggested to set $n_t = -2^j$ (see [14] and [15] for more details). Moreover, we must satisfy at least the condition $U_{0,k}(n_t) - L_{0,k}(n_t) \leq \epsilon$, returning the perfect simulation $S = (L_0(n_t) + U_0(n_t))/2$ with accuracy ϵ , where $\epsilon > 0$, and it is a “user-specified parameter.”

2) *M–H Perfect Simulation*: On the other hand, we present a generalization of the perfect M–H algorithm, which is used when the Gibbs perfect sampling from the target distribution $\varphi(\cdot)$ is impossible to do. In this case, we consider that

$$\mathcal{L}(\mathbf{y}|\theta) \propto \alpha(\theta) \prod_{k=1}^p q(\theta_k)$$

where $q(\cdot)$ is the candidate generating density, and $\alpha(\theta)$ is the probability of move $\alpha(\theta) = \alpha(\theta_{t-1,k}, \theta_{t,k})$.

- 1) The first proposal $D_{t,k}$ is generated from $q(\cdot)$ together with $R_{t,k} \sim \mathcal{U}(0, 1)$.
- 2) Second, set

$$\theta_{t,k} = \begin{cases} D_{t,k}, & \text{if } R_{t,k} \leq \min \{\alpha_k \{D_{t,k}, \theta_{t-1,k}, \theta_{t,k \leftarrow}\}, 1\} \\ \theta_{t-1,k}, & \text{otherwise} \end{cases}$$

where

$$\alpha_k \{d_k, \theta_k, \theta_{-k}\} = \frac{\varphi(\theta_1, \dots, \theta_{k-1}, d_k, \theta_{k+1}, \dots, \theta_p) q(\theta_k)}{\varphi(\theta_1, \dots, \theta_{k-1}, \theta_k, \theta_{k+1}, \dots, \theta_p) q(d_k)}$$

which is larger than 1 if $d_k \leq \theta_k$, whereas it decreases from 1 to 0 as a function of θ_{-k} when $d_k > \theta_k$.

Then, the CFTP algorithm is described as follows. At each integer n_t , put

$$\begin{aligned} L_n(n_t) &= 0 \quad t > n_t \\ L_{t,k}(n_t) &= \begin{cases} D_{t,k}, & \text{if } R_{t,k} \leq \alpha_k \{D_{t,k}, L_{t-1,k}(n_t), U_{t,k \leftarrow}(n_t)\} \\ L_{t-1,k}(n_t), & \text{otherwise} \end{cases} \end{aligned} \quad (14)$$

and

$$U_n(n_t) = D(n_t)$$

$$U_{t,k}(n_t) = \begin{cases} D_{t,k}, & \text{if } R_{t,k} \leq \alpha_k \{D_{t,k}, U_{t-1,k}(n_t) \\ & L_{t,k+}(n_t)\} \\ U_{t-1,k}(n_t), & \text{otherwise.} \end{cases} \quad (15)$$

The coalescence time is almost surely finite, satisfying the condition $U_{0,k}(n_t) - L_{0,k}(n_t) \leq \epsilon$, and the output $S = (L_0(n_t) + U_0(n_t))/2$ follows the target distribution $\varphi(\cdot)$.

V. MCMC ESTIMATION WITH A MEASURAND APPLICATION

The measurand-estimation problem presented in [21] and [17] is considered here again, as an example of a real problem. The RFEC inspection technique is used for dimensioning grooves that may occur in ferromagnetic conductive pipes. This task involves the depth and the length estimation of corrosion grooves from measurands of a picked-up coil signal phase y_i at different positions x_i closest to the defect. A finite-element calculation was performed in [17] to obtain a parametric model of the physical phenomenon. The dimensions of the groove to be estimated are linked to the parametric model through a polynomial function [see (17)]. The previous knowledge of the relationship between the groove parameters (length and depth) and the observed data (detector phase) is needed. A finite-element modeling has been used to obtain such a relation. In order to make statistical inferences about parameters and groove estimators, first, we propose the use of MCMC techniques. Secondly, we compare the obtained MCMC results with those obtained by using a primitive Monte Carlo (PMC) scheme. In this last scheme, we choose the hypothesis of errors normality ($e_i \sim \mathcal{N}(0, \hat{\sigma}_{ML}^2)$), where $\hat{\sigma}_{ML}^2$ is an unbiased variance estimate. The PMC scheme builds a likelihood empirical population by using the nonlinear least squares (NLS) estimator. In both cases, we suppose a modeling error plus acquisition error (e.g., errors between the observed data and the nonlinear model) whose distribution is assumed as Gaussian with unknown variance.

Fig. 2 illustrates the typical experimental apparatus used for groove dimensioning; the sensor is pushed inside the pipe and along with the coil position x , the phase of the detector voltage y is acquired. The distance ($2L$) between the exciter and the detector coils is chosen so that the remote field condition is satisfied.

Taking into account the symmetry and the range of the data, several mathematical functions (e.g., nonlinear models) have been considered in [17] and [21] to approximate y . The nonlinear model structure retained in [21] [f_2 ($p = 3$)] as the best model is considered here

$$f(x, \theta) = \theta_1(\arctan(\theta_2(x + \theta_3 + L)) - \arctan(\theta_2(x - \theta_3 + L)) \\ + \arctan(\theta_2(x + \theta_3 - L)) - \arctan(\theta_2(x - \theta_3 - L))) \quad (16)$$

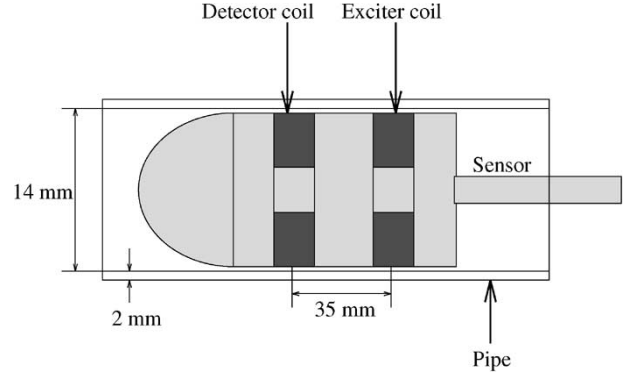


Fig. 2. Schematic illustration of the sensor pipe in the (RFEC) inspection problem.

where L is an experimental constant (half distance between coils equal to 17.5 mm), the length of vectors x , and y is $n = 118$. Fig. 3 shows an example of simulated observed data (circles) computed by a finite-element code for given groove parameters.

The measurement goal is to estimate the size of the defect [depth (d) and length (l)] from the knowledge of detector phase. The measurand quantity can be expressed as a function of the optimal parameters of the model. An algebraic form for g is chosen; g can be taken as a bilinear polynomial function of θ , which may be written as follows:

$$\hat{d} = g_d(\theta) = \sum_{k,h} c_{kh}^d \theta_k \theta_h + \sum_k c_k^d \theta_k$$

$$\hat{l} = g_l(\theta) = \sum_{k,h} c_{kh}^l \theta_k \theta_h + \sum_k c_k^l \theta_k. \quad (17)$$

The coefficients (c_{kh}^d, c_k^d) and (c_{kh}^l, c_k^l) have been computed in the least squares sense [17], for several real defect dimensions. Thus, for the candidate model, an analytical form for the measurand vector $\hat{m} = [\hat{d}, \hat{l}]^T$ as a function of θ is obtained.

A. Approximation by a Gibbs Sampling Scheme

For the MCMC approximation, we use firstly a simple Gibbs sampler in a Bayesian framework. Since errors and parameter densities are unknown, and starting off observing from the empirical errors histogram (see Fig. 4), we will make some assumptions. The assumptions of complete ignorance of parameter density are reflected by the uniform distribution. Nevertheless, Fig. 4 shows that error density tends to be of zero mean. At the moment, we could assume that errors are normally distributed (classical assumptions) with unknown variance; we also assume that the complete conditional distribution of σ^2 is inverse gamma $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$. Thus, in this case, the Bayesian framework is equivalent to the ML, and also to the least squares principle. To evaluate the estimation performance of m , its probability density has been analyzed. The parameters and measurand pdf estimation is determined by MCMC asymptotically. We have taken as reference the second example given in [8] and more general results given in [23] for autoregressive moving average (ARMA) models.

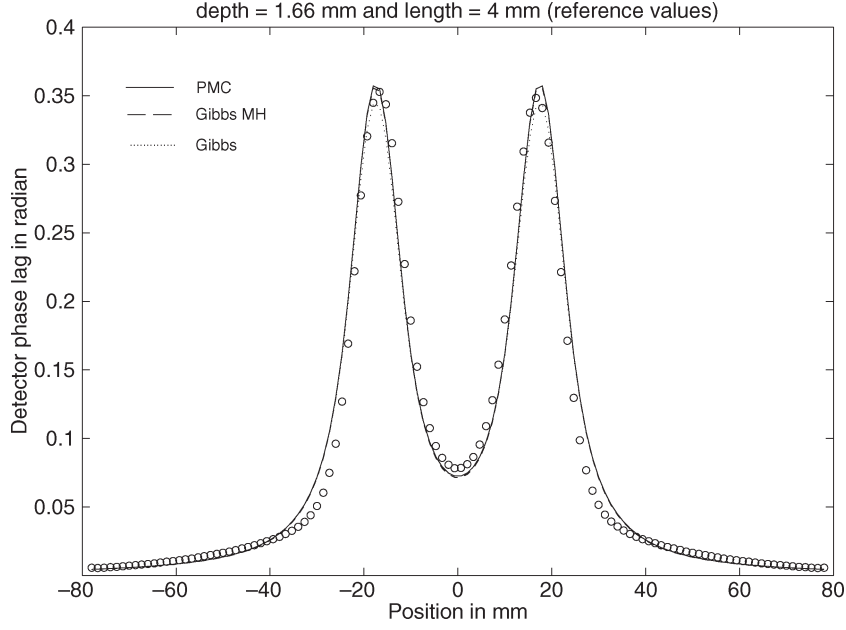


Fig. 3. Full line is the data fitted by the model using the mean parameter estimates from PMC-NLS, (...) from Gibbs and (---) from Gibbs M-H versus the simulated data (circles).

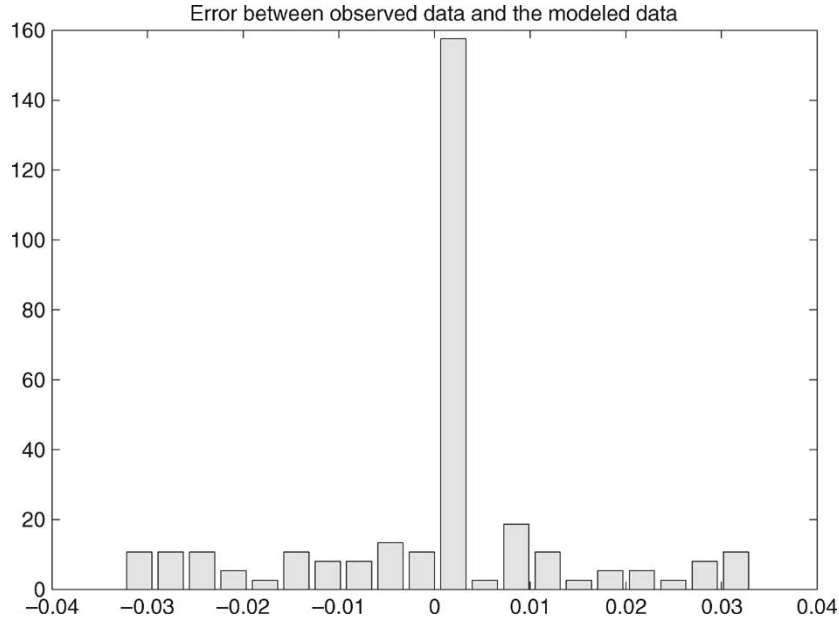


Fig. 4. Error vector obtained by $\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\theta}}_{\text{NLS}})$.

In a first normal corrected approach, the likelihood function given $\{y_i\}_{i=1}^n$ ($n = 118$) samples is the following:

$$\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi^n |\boldsymbol{\Sigma}|}} \times \exp \left[-\frac{1}{2} (\mathbf{y} - f(\mathbf{x}, \boldsymbol{\theta}))^\top \boldsymbol{\Sigma}^{-1} (\mathbf{y} - f(\mathbf{x}, \boldsymbol{\theta})) \right] \quad (18)$$

where $\boldsymbol{\Sigma} = \sigma^2 \mathbb{I}$, and the *a priori* distribution is given by the conditional distribution

$$\wp(\boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta})\pi(\sigma^2) \quad (19)$$

where $\pi(\boldsymbol{\theta}) = 1$ (e.g., parameters are assumed to be uniformly distributed), and where the complete conditional distribution of σ^2 is inverse gamma $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$, with

$$\alpha = \frac{n+p}{2}$$

and

$$\beta = \frac{1}{2} \sum_{i=1}^n (y_i - f(x_i, \boldsymbol{\theta}))^2.$$

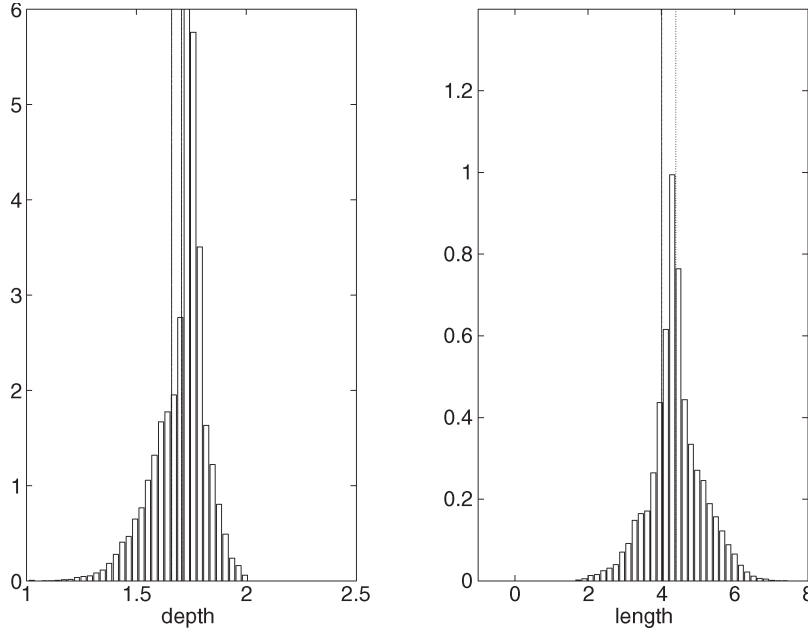


Fig. 5. Histograms of the posterior measurand pdfs for 5×10^4 PMC-NLS iterations, reference values (—), and mean values (···). The axis y indicates the nonnormalized empirical probability.

The parameter estimator is given by

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \mathcal{J}(\theta) \quad (20)$$

where $\mathcal{J}(\theta) = \sum -\log \varphi(e)$ (ML), so the likelihood is simulated from $\mathcal{L}(\mathbf{y}|\theta) = \mathcal{N}(\theta|\hat{\theta} + \mathbf{b}(\hat{\theta}), \mathbf{F}(\hat{\theta})^{-1})$ assuming the stationarity and invertibility of the nonlinear regression model, and the covariance matrix given by the inverse of the Fisher information matrix

$$\mathbf{F}(\hat{\theta})^{-1} = \left[\frac{1}{\sigma^2} \sum_{i=1}^n \frac{\partial f(x_i, \hat{\theta})}{\partial \hat{\theta}} \frac{\partial f(x_i, \hat{\theta})}{\partial \hat{\theta}^\top} \right]^{-1} \quad (21)$$

and bias \mathbf{b} is approximated using (22), also given in [24]

$$\mathbf{b}(\hat{\theta}) = -\frac{1}{2} \mathbf{F}(\hat{\theta})^{-1} \sum_{i=1}^n \frac{1}{\sigma^2} \frac{\partial f(x_i, \hat{\theta})}{\partial \hat{\theta}} \times \text{tr} \left\{ \mathbf{F}(\hat{\theta})^{-1} \frac{\partial^2 f(x_i, \hat{\theta})}{\partial \hat{\theta} \partial \hat{\theta}^\top} \right\}. \quad (22)$$

In this case, the N simulated samples are obtained from $\pi(\sigma^2, \theta|\mathbf{y})$, which is based on the full conditional densities of σ^2 and θ given by

- 1) the density of σ^2 given θ and \mathbf{y} , $\pi(\sigma^2|\mathbf{y}, \theta)$;
- 2) density of θ given σ^2 and \mathbf{y} , $\pi(\theta|\mathbf{y}, \sigma^2)$.

The posterior probability is given by $\varphi(\theta|\mathbf{y}) \propto \mathcal{L}(\mathbf{y}|\theta)$, which is obtained sampling from its full conditional density $\pi(\sigma^2, \theta|\mathbf{y})$ (e.g., Gibbs sampler), then one calculates the nonlinear mapping $g_\ell(\theta)$ at each sample, which is straightforward, obtaining the posterior probability density $\varphi(m_\ell|\theta)$. Results obtained by a PMC procedure are compared versus Gibbs single chain (MCMC¹) for 6×10^3 replications [the first 1000 replications have been ignored (burn in), collecting the

next 5000] and Gibbs perfect sampling (MCMC², double chain) for 4500 replications (the first 500 replications have been ignored, collecting the next 4000). The results obtained after ten MCMC independent simulations of a single chain and perfect sampling has shown that there is almost convergence to the parameters and measurand pdfs.

Fig. 3 illustrates the data fitted using the mean parameter estimates (pointed line) from the Gibbs sampler. In this case, we have used the noncentered weighted bootstrap for convergence control. Figs. 5 and 6 also show the difference between measurand populations with the Gibbs perfect-sampling scheme (MCMC²) and the PMC-NLS scheme. In the comparison between these figures, it is clear that the Gibbs approach is in the measurand population neighborhood (see also Table I), but the nonlinearity of parameters and that one of the measurand is not taken into account by the normal approximation of $\mathcal{L}(\mathbf{y}|\theta)$. The Pázman approach could be a better alternative to describe these nonlinearities, also with normal error assumptions. However, in such a case, we need to use the M-H algorithm. Finally, Table I presents some statistics approximated by both Gibbs schemes (resampling and perfect-simulation algorithm); the nonlinearity of θ_2 is remarkable by the large difference between the mean and the median values obtained by PMC-NLS and MCMC. The PMC-NLS simulation mean time is 30 times more expensive than both Gibbs normal approaches to obtain the approximated statistics shown in Table I. Moreover, if the nonlinearity of the model $f(\cdot)$ decreases, the Gibbs approaches proposed here will give a better approximation with the quickest convergence.

B. Approximation by a Gibbs M-H Scheme

The second MCMC approach is derived by changing the likelihood function; this function will be approximated by the deterministic Pázman parameter density. Our first interest is

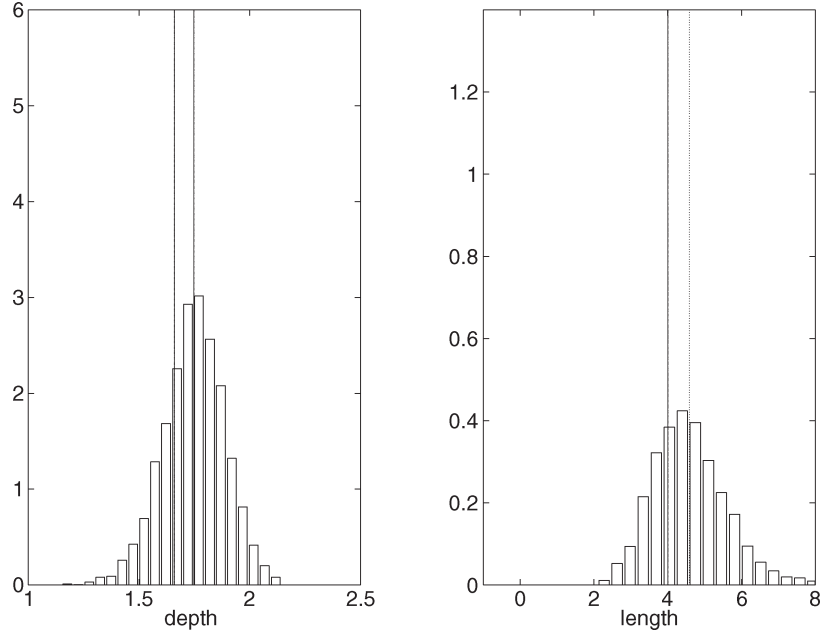


Fig. 6. Histograms of the posterior measurand pdfs for 6×10^3 Gibbs (MCMC²) iterations, reference values (—), and mean values (···). The axis y indicates the nonnormalized empirical probability.

TABLE I
STATISTICS OBTAINED BY THE GIBBS SAMPLER WITH RESAMPLING (MCMC¹) AND PERFECT-SIMULATION ALGORITHMS (MCMC²) VERSUS PMC-NLS FOR θ AND THE MEASURAND

PMC-NLS	True	Mean	Median	Bias	Var	std
θ_1	unknown	0.3097	0.3078	0.0148	0.0017	0.0409
θ_2	unknown	0.1825	0.1815	1.2e-3	4.0761e-5	0.0064
θ_3	unknown	3.4694	3.4584	0.1075	0.1687	0.4107
d	1.6600	1.7048	1.7292	0.0448	0.0132	0.1149
l	4.0000	4.3915	4.3574	0.3915	0.5153	0.7179
MCMC ¹						
θ_1	unknown	0.3001	0.2992	0.0052	0.0024	0.0491
θ_2	unknown	0.1783	0.1784	0.0054	6.7908e-5	0.0082
θ_3	unknown	3.5140	3.5245	0.0629	0.3303	0.5747
d	1.6600	1.7499	1.7534	0.0899	0.0197	0.1402
l	4.0000	4.5707	4.4967	0.5754	0.9853	0.9926
MCMC ²						
θ_1	unknown	0.2991	0.2982	0.0042	0.0028	0.0526
θ_2	unknown	0.1787	0.1788	0.0050	7.7830e-5	0.0088
θ_3	unknown	3.5256	3.5344	0.0513	0.3796	0.6161
d	1.6600	1.7436	1.7483	0.0174	0.0227	0.1509
l	4.0000	4.6035	4.5226	0.5568	1.1421	1.0704

sampling from the likelihood pdf, extracting samples from the whole joint probability density; second, we use these samples to complete the posterior sampling. Pázman [25] proposes a differential geometrical approach, which leads to an estimation of the whole parameter pdf under standard assumptions (Gaussian error assumptions). Considering a nonasymptotical case (finite sample n), Pázman has given a formula to approximate the pdf of the classical least squares parameter estimate. The formula is expressed in terms containing zeroth-, first-, and second-order

derivatives of $f(\cdot)$; it is computed at the point $\hat{\theta}$, since $\bar{\theta}$ is unknown. The whole pdf evaluation is driven by (e.g., the joint probability density)

$$\hat{\varphi}(\mathbf{y}|\hat{\theta}) = \frac{\det \mathbf{Q}(\hat{\theta})}{(2\pi)^{\frac{p}{2}} \det^{\frac{1}{2}} \mathbf{F}(\hat{\theta})} \times \exp \left\{ -\frac{1}{2} \left\| \mathbf{P}(\hat{\theta}) \left[f(\mathbf{x}, \hat{\theta}) - \bar{f} \right] \right\|_{\Sigma}^2 \right\} \quad (23)$$

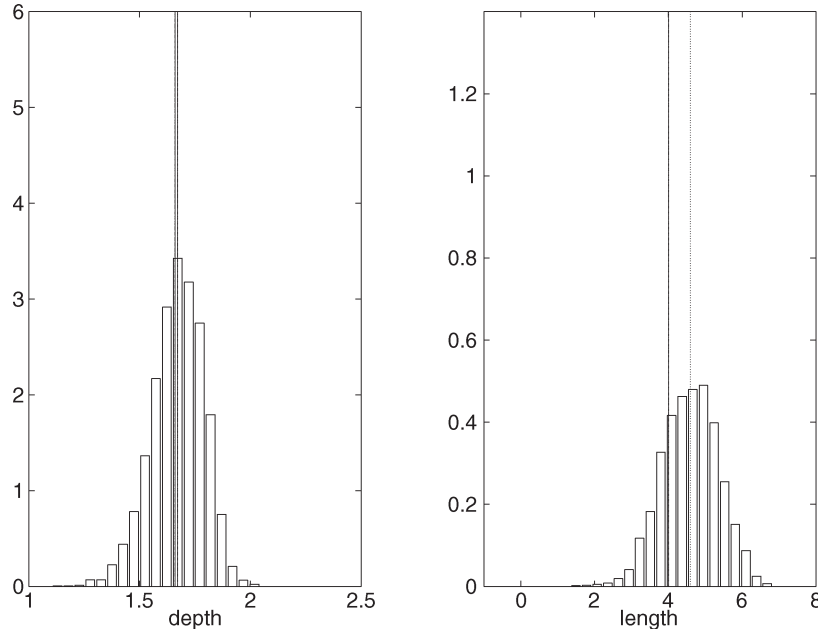


Fig. 7. Histograms of the posterior measurand pdfs for 4500 Gibbs M-H (G-MH²) iterations, reference values (—), and mean values (···). The axis y indicates the nonnormalized empirical probability.

where \bar{f} is the true mean of \mathbf{y} , $\mathbf{F}(\boldsymbol{\theta})$ is the $(p \times p)$ Fisher information matrix [see (21)], with $\boldsymbol{\Sigma}^{-1} = 1/\sigma^2 \mathbb{I}$ ($n \times n$) matrix for a given error variance σ^2 , $\mathbf{P}(\boldsymbol{\theta})$ is the $(n \times n)$ projection matrix

$$\mathbf{P}(\boldsymbol{\theta}) = \sum_{i,j=1}^p \frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i} \mathbf{F}_{ij}^{-1}(\boldsymbol{\theta}) \frac{\partial f^T(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_j} \boldsymbol{\Sigma}^{-1}$$

and $\mathbf{Q}(\boldsymbol{\theta})$ is a $(p \times p)$ matrix

$$\mathbf{Q}_{ij}(\boldsymbol{\theta}) = \mathbf{F}_{ij}(\boldsymbol{\theta}) + (f(\mathbf{x}, \boldsymbol{\theta}) - \bar{f}) (\mathbb{I} - \mathbf{P}(\boldsymbol{\theta})) \frac{\partial^2 f(\mathbf{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}.$$

As the formula suggests, the $\boldsymbol{\theta} \in \Theta$ has a support that exists and could be defined as a local support if the “real” parameters are known. The target density used for the M-H algorithm is then $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}) = \hat{\varphi}(\mathbf{y}|\boldsymbol{\theta})$ according to (23). Multi-dimensional integration and support searching are driven by the M-H autoregressive algorithm, the performance of which is generally better than the M-H random walk [8].

An autoregressive chain is represented by a vector autoregressive process of order 1. This algorithm is due to the following procedure of candidate production

$$\boldsymbol{\theta}_{(t+1)} = \boldsymbol{\theta}_0 + \mathbf{B}(\boldsymbol{\theta}_t - \boldsymbol{\theta}_0) + \mathbf{z} \quad (24)$$

where $\boldsymbol{\theta}_0$ is a vector and \mathbf{B} is a matrix, both can be confirmed with $\boldsymbol{\theta}_t$, and \mathbf{z} has $p_2(\cdot)$ as its density. Then, the transition kernel is $q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{(t+1)}) = p_2(\boldsymbol{\theta}_{(t+1)} - \boldsymbol{\theta}_0 - \mathbf{B}(\boldsymbol{\theta}_t - \boldsymbol{\theta}_0))$. This autoregressive version is an intermediate between independent ($\mathbf{B} = 0$) and random-walk ($\mathbf{B} = \mathbb{I}$) versions. If $\mathbf{B} = -\mathbb{I}$, this algorithm produces chains that are reflected about the point $\boldsymbol{\theta}_0$ and is a simple way to induce negative correlation between successive elements of the chain ($\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_{(t+1)}$), and it permits a quick scanning of the surface of $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})$. The algorithm is similar to the M-H basic algorithm, where $q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{(t+1)}) = p_2(\cdot)$,

and \mathbf{z} is a random vector drawn by sampling from the uniform distribution $\mathbf{z} \sim \mathcal{U}(0, \Delta)$, with $\Delta = \text{diag}(\delta_1, \dots, \delta_p)$.

The basic M-H algorithm is summarized as follows.

- 1) The initial value of $\boldsymbol{\theta}_0$ is given by the real parameter value or the nonlinear least squares estimate.
- 2) For $t = 1, \dots, N$, repeat.
- 3) Generate $\boldsymbol{\theta}_{(t+1)} \sim q(\boldsymbol{\theta}_t, \cdot)$ [see (24)] and $u \sim \mathcal{U}(0, 1)$.
- 4) If $u \leq \alpha(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{(t+1)})$, where

$$\alpha(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{(t+1)}) = \begin{cases} \min \left\{ \frac{\pi(\boldsymbol{\theta}_{(t+1)})q(\boldsymbol{\theta}_{(t+1)}, \boldsymbol{\theta}_t)}{\pi(\boldsymbol{\theta}_t)q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{(t+1)})}, 1 \right\}, & \text{if } \pi(\boldsymbol{\theta}_t)q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{(t+1)}) > 0 \\ 1, & \text{otherwise.} \end{cases}$$

Then, set $\boldsymbol{\theta}_{(t+1)} = \boldsymbol{\theta}_{(t+1)}$.

- 5) Else, set $\boldsymbol{\theta}_{(t+1)} = \boldsymbol{\theta}_t$.
- 6) Return the values $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_N\}$.

The initial value of $\boldsymbol{\theta}$ is the “real” parameter vector $\boldsymbol{\theta}_0$ (unknown in practice) or its ML estimate. The Δ matrix can also be given by $\lambda \mathbf{K}(\boldsymbol{\theta})$, where $\mathbf{K}(\boldsymbol{\theta})$ is obtained by the Cholesky decomposition of $\mathbf{F}(\boldsymbol{\theta})^{-1}$, where $\lambda = 3.2$; it was chosen heuristically. The *a priori* distribution $\varphi(\boldsymbol{\theta})$ is given as in the normal likelihood approach (Section V-A), and the rest of the algorithm is similar.

The PMC-NLS results are also compared versus Gibbs M-H single chain (G-MH¹) for 6000 replications and versus Gibbs M-H perfect sampling (G-MH² double chain) for 4500 replications. Fig. 3 illustrates the data fitted using the mean parameter estimates (dashed line) from the Gibbs M-H scheme. In this case, we compare results obtained by Gibbs perfect sampling and Gibbs M-H perfect sampling versus PMC-NLS results. Figs. 5 and 7 show also the difference between the measurand populations with the Gibbs M-H perfect-sampling scheme and the PMC-NLS scheme. Finally, it is clear that the Gibbs M-H

TABLE II
STATISTICS OBTAINED BY THE GIBBS M–H SAMPLER WITH SINGLE CHAIN (G-M-H¹) AND
PERFECT SIMULATION (G-M-H²) ALGORITHMS FOR θ AND THE MEASURAND

G-M-H ¹	True	Mean	Median	Bias	Var	std
θ_1	unknown	0.3103	0.3010	0.0154	0.0024	0.0494
θ_2	unknown	0.1834	0.1822	3.0e-4	5.0648e-5	0.0071
θ_3	unknown	3.4752	3.5049	0.1017	0.2327	0.4824
d	1.6600	1.6917	1.7156	0.0317	0.0173	0.1316
l	4.0000	4.4100	4.4516	0.4100	0.7123	0.8440
G-M-H ²						
θ_1	unknown	0.2979	0.2924	0.0030	0.0017	0.0407
θ_2	unknown	0.1846	0.1844	9.0e-4	4.1251e-5	0.0064
θ_3	unknown	3.5832	3.5925	0.0063	0.1849	0.4300
d	1.6600	1.6684	1.6770	0.0084	0.0139	0.1180
l	4.0000	4.6070	4.6226	0.6070	0.5738	0.7575

TABLE III
MEAN ESTIMATES OBTAINED FOR d AND l BY THE PMC-NLS AND THE MCMC SCHEMES AND BOTH MSE MEASURES

\	PMC-NLS	MCMC ¹	MCMC ²	G-M-H ¹	G-M-H ²
MSE $_{\theta}$	1.9332e-4	1.9192e-4	1.9147e-4	1.9715e-4	1.9124e-4
\bar{d}	1.7048	1.7499	1.7436	1.6917	1.6684
\bar{l}	4.3915	4.5707	4.6035	4.4100	4.6070
MSE $_m$	0.0776	0.1669	0.1856	0.0846	0.1843

approach is close to the measurand population neighborhood of PMC-NLS (previously seen in Fig. 5), and the nonlinearity of parameters and the measurand is best described than by the normal approximation of $\mathcal{L}(y|\theta)$ (see Fig. 6).

The percentage of acceptance of the Gibbs M–H single chain diminishes to 18.35%, obtaining 81.65% of rejected samples, which assures almost convergence of this scheme. The PMC-NLS simulation mean time is six times more expensive than the Gibbs M–H perfect-sampling simulation procedure to obtain the approximated statistics shown in Table II. The simulation time is also a great issue. For example, for the 5×10^4 PMC iterations of the NLS scheme for the model used for RFEC application, the time needed to achieve convergence of simulation was 7.32 h, whereas the Gibbs M–H simulation time was in the interval of 72.07–46.67 min when using an ultra Sparc 5 system (256-MB RAM), and MATLAB version 6. On the other hand, the level of approximation not only depends on the model $f(\cdot)$ but also on the polynomial (nonlinear) measurand mapping $g_{\ell}(\cdot)$.

Finally, we have taken advantage from the mean parameter and measurand estimates to obtain the mean-squared error (MSE) from the different schemes studied here, with the purpose of comparing the MSE in the fitting sense: $\text{MSE}_{\theta} = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i, \hat{\theta}))^2$ and the MSE in the measurand sense: $\text{MSE}_m = \frac{1}{2} \sum_{j=1}^2 (m_{rj} - \hat{m}_j)^2$. The best parameter mean values are those estimated with the G-MH² sampling scheme (see Table III), since the MSE in the fitting sense is the smallest. On the other hand, the best measurand mean values are those obtained by the PMC-NLS scheme. In the case of this

application, the normal error assumptions could be verified by means of the analysis of its CDF or other tests.

VI. CONCLUDING REMARKS

Generally, the MCMC methods offer an easy pdf estimation in a Bayesian framework. The nonlinearity of the parametric model and the measurand is best described by using the Pázman approach as likelihood target joint pdf with respect to the normal approximation (i.e., Gibbs). Both MCMC Bayesian proposed approaches give a good statistical approximation, and since the Bayesian framework presented in this case is equivalent to the ML, both are comparable. Nevertheless, in the Bayesian framework, it is supposed that the error variance follows a conditional pdf, which is not the case in the PMC approach, where the error variance must be well specified with a fixed value (one can use a good variance-estimate version). If the nonlinearity of the model $f(\cdot)$ diminishes, then the normal likelihood approximation gives good results with the quickest convergence. Moreover, we can preserve the same MCMC schemes when the *a priori* distribution on parameters is other than the noninformative *a priori* one. The current work considers non-Gaussian assumptions of errors, where the Gibbs sampler seems to be more complex, and M–H could help to find the problem solution; here also, one can postulate hybrid MCMC algorithms. The weighted resampling and acceptance–rejection percentage methods are nowadays well understood; however, perfect-sampling methods are recent and interesting tools, which could be exploited in a large class of problems [16], since the convergence improvement using CFTP

could be reached. Finally, some recent research works point toward the use of slice sampling methods [26]; thus, we propose a study of these methods in order to apply this new idea for future works on the measurand-estimation problem.

ACKNOWLEDGMENT

Many thanks to reviewers for the diligent review and useful comments given, which improved the present manuscript.

REFERENCES

- [1] G. Fleury, "Optimal nonlinear modeling and reparametrization," in *Proc. IEEE Int. Workshop Intelligent Signal Processing*, Budapest, Hungary, Sep. 1999, pp. 72–76.
- [2] J. Besag, "Markov Chain Monte Carlo for statistical inference," Univ. Washington, Center Statistics Social Sci., Seattle, WA, Tech. Rep. 9, 2000.
- [3] G. S. Fishman, "Monte Carlo concepts, algorithms, and applications," in *Springer Series in Operations Research*, 3rd ed. New York: Springer-Verlag, 1995–1999.
- [4] R. M. Neal, "Probabilistic inference using Markov Chain Monte Carlo methods," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. CRG-TR-93-1, 1993. [Online]. Available: www.cs.toronto.edu/~radford
- [5] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York: Springer-Verlag, 2004.
- [6] B. P. Carlin and S. Chib, "Bayesian model choice via Markov Chain Monte Carlo methods," *J. R. Stat. Soc., Ser. B*, vol. 57, no. 3, pp. 473–484, 1995.
- [7] A. F. M. Smith and G. O. Roberts, "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods," *J. R. Stat. Soc., Ser. B*, vol. 55, no. 1, pp. 3–23, 1993.
- [8] S. Chib and E. Greenberg, "Understating the Metropolis–Hastings algorithm," *Amer. Stat.*, vol. 49, no. 4, pp. 327–335, Nov. 1995.
- [9] S. Senecal and P.-O. Amblard, "MCMC methods for discrete source separation," in *Proc. MaxEnt, Bayesian Inference and Maximum Entropy Methods Science Engineering*, A. Mohammad-Djafari, Ed, Gif-sur-Yvette, France, Jul. 2000, pp. 350–359.
- [10] J. Vermaak, M. Niranjan, and S. J. Godsill, "An improved speech production model for voiced speech utilising a seasonal AR-AR model and MCMC simulation," Signal Processing Group, Cambridge Univ., Cambridge, U.K., Tech. Rep. CUED/F-INFENG/TR.325, 1998. [Online]. Available: www.statslab.cam.ac.uk/~mcmc/pages/list.html
- [11] A. E. Gelfand and S. K. Sahu, "On Markov Chain Monte Carlo acceleration," *J. Comput. Graph. Stat.*, vol. 3, no. 3, pp. 261–276, 1994.
- [12] P. Hall and B. Presnell, "Intentionally biased Bootstrap methods," *J. R. Stat. Soc., Ser. B*, vol. 61, no. 1, pp. 143–158, 1999.
- [13] J. Shao, "Bootstrap model selection," *J. Amer. Stat. Assoc.*, vol. 91, no. 434, pp. 655–665, Jun. 1996.
- [14] J. Möller, "Perfect simulation of conditionally specified models," *J. R. Stat. Soc., Ser. B*, vol. 61, no. 1, pp. 251–264, 1999.
- [15] J. G. Propp and D. B. Wilson, "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Struct. Algorithms*, vol. 9, no. 1/2, pp. 223–252, 1996.
- [16] P. Djurić, Y. Huang, and T. Ghirmai, "Perfect sampling: A review and applications to signal processing," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 345–356, Feb. 2002.
- [17] M.-E. Davoust, G. Fleury, and J. Oksman, "A parametric estimation approach for grooves dimensioning using remote field eddy current inspection," *Res. Nondestruct. Eval.*, vol. 11, no. 1, pp. 39–57, 1999.
- [18] G. Fleury, "Non-intrusive time-of-flight flowmeter-parametric estimation and optimization," *Sens. Actuators A, Phys.*, vol. 47, no. 1, pp. 364–368, Mar. 1995.
- [19] J. I. De la Rosa and G. Fleury, "Bootstrap density approximation for measurement purposes," in *Proc. Int. Association Science and Technology Development (IASTED) Conf. Signal Processing and Communications*, Marbella, Spain, Sep. 2000, pp. 357–361.
- [20] —, "Bootstrap methods applied to indirect measurement," in *Proc. Physics Signal and Image Processing Conf. (PSIP)*, Marseille, France, Jan. 2001, pp. 57–62.
- [21] S. Brahimi-Belhouari, G. Fleury, and M. E. Davoust, "Model selection based on robustness criterion with measurement application," in *Proc. SPIE*, Denver, CO, 1999, vol. 3816, pp. 196–202.
- [22] S. Brahimi-Belhouari, M. Kieffer, G. Fleury, L. Jaulin, and E. Walter, "Model selection via worst-case criterion for nonlinear bounded-error estimation," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 3, pp. 653–658, Jun. 2000.
- [23] J. Marriott, N. Ravishanker, A. Gelfand, and J. Pai, "Bayesian analysis of ARMA processes: Complete sampling-based inference under exact likelihoods," in *Bayesian Analysis in Statistics and Econometrics*, D. A. Berry, K. M. Chaloner, and J. K. Geweke, Eds. New York: Wiley, 1996, pp. 243–256.
- [24] M. J. Box, "Bias in nonlinear estimation," *J. R. Stat. Soc., Ser. B*, vol. 33, no. 1, pp. 171–201, 1971.
- [25] A. Pázman, "Probability distribution of the multivariate nonlinear least squares estimates," *Kibernetika (Prague)*, vol. 20, no. 2, pp. 209–230, 1984.
- [26] R. M. Neal, "Slice sampling," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. 2005, 2001.



José I. De la Rosa (S'01–A'02–M'04) was born in Zacatecas, Mexico, on March 19, 1972. He received the B.Sc. degree in electronics and telecommunications engineering from the Universidad Autónoma de Zacatecas, in 1995, the M.Sc. degree in digital systems from the Instituto Politécnico Nacional (CITEDI), Mexico City, Mexico, in 1998, and the Ph.D. degree in signal processing and control from the Université de Paris-Sud, Orsay, France, in 2002. He was pursued the Ph.D. degree with the Measurement Department (now Department of Signal Processing and Electronic Systems) at École Supérieure d'Électricité (SUPELEC).

He is presently an Associate Researcher at the Signal Processing Laboratory of the engineering faculty from the Universidad Autónoma de Zacatecas. His main interests are statistical signal processing, modeling and system identification, and signal and image processing in robotics.



Gilles A. Fleury was born in Bordeaux, France, on January 8, 1968. He received the B.Sc. degree in electronics and signal processing from the École Supérieure d'Électricité (SUPELEC), Gif-sur-Yvette France, in 1990, the Ph.D. degree in signal processing from the Université de Paris-Sud, Orsay, France, in 1994, and the HDR degree in 2003.

He is presently a Professor with the Department of Signal Processing and Electronic Systems, SUPELEC. He has worked in the areas of inverse problems and optimal design. His current research interests include bioinformatics, optimal nonlinear modeling, and nonuniform sampling.



Sonia E. Osuna was born in Sonora, Mexico, in 1972. She received the B.Sc. degree in electronics and computing engineering from the Instituto Tecnológico de Sonora in 1998, the M.Sc. degree in digital systems from the Instituto Politécnico Nacional (CITEDI), Mexico City, Mexico, in 2003. She pursued the M.Sc. degree within the Measurement Department, SUPELEC, Gif-sur-Yvette, France, in 2001.

Her main research interests are statistical signal processing and telecommunications.



Marie-Eve Davoust was born in Thorigny-sur-Marne, France, in 1948. She received the Ph.D. degree in physics from the University of Paris 6, Paris, France, in 1978.

She is currently a Professor at SUPELEC, Gif-sur-Yvette, France. She has been involved in various research topics in signal processing. Her main interests are physical system modeling and signal processing for measurement purposes and microsystems computer-aided design. She teaches courses on sensors and nondestructive evaluation.